Executive function in children with externalizing and comorbid internalizing behavior problems

Steven Woltering,¹,² Victoria Lishak,² Nick Hodgson,² Isabela Granic,³ and Philip David Zelazo⁴

¹Educational Psychology, Texas A&M University, College Station, TX, USA; ²Applied Psychology and Human Development, Ontario Institute for Studies in Education, University of Toronto, Toronto, ON, Canada; ³Developmental Psychopathology Department, Radboud University Nijmegen, Nijmegen, The Netherlands; ⁴College of Education and Human Development, University of Minnesota, Minneapolis, MN, USA

Background: The goal of this study is to investigate differences in executive function (EF) in children with different levels of disruptive behavior problems (DBP). Methods: Ninety-three children between 7 and 12 years old with DBP were compared to 63 normally developing peers on a battery of EF tasks that varied in the amount of required emotion regulation ('hot' EF). Results: Differences in EF were found between DBP and comparison groups as indexed by hot EF tasks. Self-reported emotion scales, in conjunction with physiological recordings of heart rate, confirmed that emotions were elicited during hot EF. Conclusions: Results suggest that difficulties in hot EF underlie externalizing problem behaviors in middle childhood.

Keywords: Executive function, emotion regulation.
The distinction between pure externalizing and anxiety comorbid samples may also have consequences for understanding EF during different emotion contexts. For example, callous unemotional traits have been associated with a subset of children with DBP (Frick & White, 2008; Marsh et al., 2008) and it is possible that EF skills, in particular hot EF, in these children would show less impairment compared to DBP children with comorbid anxiety. The latter group could be more susceptible to cognitively interfering effects of emotion.

The first aim of our study was to confirm the presence of EF impairments among children with DBP. We compared children with DBP to a healthy comparison group on a battery of different EF tasks as well as a single task that differed in levels of motivational or emotional significance (i.e. hot vs. cool). We hypothesized that group differences would be strongest for the hot EF tasks with poorer performance for the DBP group. To confirm the presence of emotional arousal, physiological measures were included, such as heart rate reactivity (Kreibig, 2010). Our second aim was to assess how EF would be impared across different levels of internalizing problem behaviors. Accordingly, we divided our DBP sample into one group with pure externalizing behavior problems (EXT) and another group with comorbid problems of anxiety and depression (MIXED). Although more tentative, we hypothesized that the MIXED group might be more sensitive to impairments in hot EF, given the presence of comorbidity.

Methods
Participants

The data from this analysis were taken from a larger study investigating individual differences in, and intervention effects of, the neural correlates and psychophysiological correlates of self-regulation (e.g. see Woltering, Granic, Lamm, & Lewis, 2011; Woltering, Lishak, Elliott, Ferraro, & Granic, 2015). A total of 153 children between 7 and 12 years from whom EF data were collected were included in this study. The DBP group consisted of 90 children (mean age = 9.6 years, SD = 1.8; 80% boys) and who were recruited from two children’s community mental health agencies that offered a combined parent management therapy/Cognitive Behavioral Therapy treatment for children with severe behavioral problems related to aggression within their families. Children were referred to these agencies for DBP either by a mental health professional, parent, police, or teacher. In addition, 63 typically developing children (mean age = 9.8 years, SD = 1.5; 65% boys), matched for age, were recruited through newspaper ads. The study was approved by the Research Ethics board of the University of Toronto.

The inclusion criteria for the DBP group was a score above 64 (98th percentile, clinical range) on the externalizing subscale of the Child Behavior Checklist (CBCL; Achenbach, 1991). Exclusion criteria consisted of any significant cognitive impairment, such as a persistent developmental delay. For the physiological recordings, participants needed to be free of any self-reported cardiac conditions. No participants were excluded on these grounds. Twenty-nine children were classified as EXT (mean age = 9.0 years, SD = 1.1; 79% boys) and 61 children were classified as MIXED (mean age = 9.8 years, SD = 1.1; 80% boys). The MIXED group, in addition to having high scores on the externalizing scale, also scored above 64 on the CBCL internalizing scale.

Medication use could not be regulated in this study and was therefore statistically controlled for. In the DBP group, 37 participants (41%) were using medication of which the majority (n = 29) used stimulants (remaining participants used over-the-counter medication). Medication use was similar between EXT (38%) and MIXED (43%). No children in the comparison group were on medication.

Table 1 shows the sample demographics, separately for the clinical and nonclinical group. The first group is also broken

<table>
<thead>
<tr>
<th>Demographic characteristic</th>
<th>Clinical: Extremalizers (n = 29)</th>
<th>Clinical: Mixed (n = 61)</th>
<th>Nonclinical group (n = 63)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Living arrangements (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both parents</td>
<td>8 (32)</td>
<td>21 (34)</td>
<td>32 (54)</td>
</tr>
<tr>
<td>With step-parent</td>
<td>2 (8)</td>
<td>13 (21)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Mother only</td>
<td>11 (44)</td>
<td>19 (31)</td>
<td>22 (37)</td>
</tr>
<tr>
<td>Other</td>
<td>6 (16)</td>
<td>8 (13)</td>
<td>4 (7)</td>
</tr>
<tr>
<td>Ethnicity (%)</td>
<td>21 (84)</td>
<td>45 (75)</td>
<td>27 (46)</td>
</tr>
<tr>
<td>European</td>
<td>3 (12)</td>
<td>10 (17)</td>
<td>8 (14)</td>
</tr>
<tr>
<td>African/Caribbean</td>
<td>–</td>
<td>–</td>
<td>4 (7)</td>
</tr>
<tr>
<td>Latin</td>
<td>–</td>
<td>–</td>
<td>11 (19)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (4)</td>
<td>5 (7)</td>
<td>6 (10)</td>
</tr>
<tr>
<td>Mother’s Education (%)</td>
<td>11 (44)</td>
<td>25 (41)</td>
<td>8 (14)</td>
</tr>
<tr>
<td>High school, or less</td>
<td>6 (24)</td>
<td>22 (36)</td>
<td>20 (34)</td>
</tr>
<tr>
<td>Community college</td>
<td>7 (28)</td>
<td>11 (18)</td>
<td>30 (50)</td>
</tr>
<tr>
<td>University, or higher</td>
<td>1 (4)</td>
<td>3 (5)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Father’s education (%)</td>
<td>6 (38)</td>
<td>23 (34)</td>
<td>11 (23)</td>
</tr>
<tr>
<td>High school, or less</td>
<td>7 (44)</td>
<td>12 (28)</td>
<td>9 (19)</td>
</tr>
<tr>
<td>Community college</td>
<td>3 (19)</td>
<td>6 (14)</td>
<td>27 (56)</td>
</tr>
<tr>
<td>University, or higher</td>
<td>–</td>
<td>2 (5)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Other/Unknown</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family income (CND$) (%)</td>
<td>0–29,9999</td>
<td>5 (21)</td>
<td>13 (22)</td>
</tr>
<tr>
<td>30,000–59,999</td>
<td>8 (33)</td>
<td>10 (18)</td>
<td>22 (37)</td>
</tr>
<tr>
<td>60,000 or above</td>
<td>11 (46)</td>
<td>22 (39)</td>
<td>24 (41)</td>
</tr>
</tbody>
</table>

*Mother’s/Father’s education is highest level completed.

© 2015 Association for Child and Adolescent Mental Health.
down for MIXED and EXT. Groups were different on the basis of Chi-square tests for all demographics (all p's < .01), except for income. Individually controlling for demographics did not change the overall pattern of results on the major analyses.

Procedure

Families were invited to the lab before they underwent treatment. Several cool and hot EF measures were included in the study. For cool measures, we used two inhibitory control tasks (Go–Nogo and Stroop Color-word) and a simple working memory task (Digit span; See Miyake et al., 2000, for a classification of EF tasks based on the components involved). Hot EF measures involved motivationally significant components (e.g., meaningful wins and losses). These included two decision making tasks (The Iowa Gambling Task and Delay Discounting), and a hot (emotional) version of the Go–Nogo inhibition task. The tasks were presented on a computer screen alternating between hot and cool tasks in a following order: Iowa Gambling Task (IGT), Stroop Color-word, Delay Discounting, and Digit Span. Upon completion, participants were told that they would play the Go–Nogo task and were shown two bins of toys. One bin contained small, less desirable toys such as small plastic cars, and the other bin contained larger, more desirable toys such as large action figures, stuffed animals, games, and $10 gift certificates from a local game store. Participants were asked to choose a toy from each of the bins. They were then told that if they performed successfully on the Go–Nogo task, they would receive the more desirable toy, but that if they performed poorly, they would receive only the less desirable toy. Children were then given instructions on the game and given a practice block of 30 trials.

Clinical questionnaires

Child Behavior Checklist. The CBCL (Achenbach, 1991) is a highly reliable and commonly used parent report yielding standardized T-scores for scales such as internalizing and externalizing behavior problems.

Early Adolescent Temperament Questionnaire – revised (EATQ: parent report form). The effortful control scale of the EATQ (Ellis & Rothbart, 2001) was of particular interest as it predicts mental health outcomes such as internalizing and externalizing disorders (Odloshinkel, Rosmalen, Veenstra, Dijkstra, & Orem, 2007). The effortful control scale consists of the following subscales: inhibitory control, activation control, and attention (Ellis & Rothbart, 2001; Muris, van der Pennen, Sigmond, & Mayer, 2008) and is conceptually related to self-regulation.

Executive function measures

Iowa Gambling Task. Children were administered a computerized version of the IGT (Bechara, Damasio, Damasio, & Anderson, 1994). In short, participants were shown four decks of cards (A, B, C, D) that would reveal either gains or losses (measured in play money that they were asked to treat as real). Decks A and B produced large monetary gains but even larger losses, so that, over consecutive trials, choices from these decks resulted in a substantial loss of money. Decks C and D, on the other hand, produced smaller monetary gains, but even smaller losses, resulting in an increase in money over consecutive plays. A performance score was computed by subtracting the number of disadvantageous choices from the number of advantageous choices for the last 20 trials of the task (Monterosso, Ehrman, Napier, O’Brien, & Childress, 2001). Positive difference scores indicated advantageous choices. Being a(n affective) decision making task, we will classify the IGT as a hot EF task (Kerr & Zelazo, 2004; Koechlin & Hyafil, 2007; Mitchell, 2011; Toplak, Sorge, Benoit, West, & Stanovich, 2010).

Stroop Color-word Test. The Stroop Color-word interference score (reaction time on incongruent minus congruent items) was used as a measure of performance in the analysis (Stroop, 1935). Higher scores indicated better performance (see also Nigg, 2000). The Stroop task is considered a cool EF measure of inhibitory control.

Delay Discounting. A computerized Delay Discounting task was adapted from Richards, Zhang, Mitchell, and de Wit (1999). The task was used to measure the rate at which participants discounted the value of delayed reinforcers. Participants were given a series of choices between a small immediate amount of money and $10, which would be delayed by 1, 2, 30, 180, or 365 days. For each delay period, an adjusting-amount algorithm adjusted the magnitude of the immediate reinforcer until it was equal in value to the delayed reward. The value of the adjusted immediate reinforcer was referred to as the discounting point. The measure of performance was obtained by computing k values for each participant. The k value represents the rate at which participants discounted the delayed reward in favor of the immediate reward. k values were computed by fitting participants’ data to the hyperbolic function, \( f(k) = A/(1 + kD) \) (Mazur, 1987). In this function, A is the nominal amount of the delayed reward ($10) and D is the length of the delay. Once k values were computed, they were log transformed because their distribution was skewed (Johnson & Bickel, 2002). To ensure appropriate engagement with the task, participants were told that when finished, one of their hypothetical choices would in fact be granted at random (e.g. if they were to choose 8 dollars 2 weeks from now, we said we would mail it to them in 2 weeks). For practical reasons, all children were simply given $10 upon completion of the task regardless of their choices. Lower log k values indicated better performance. Similar to the IGT, this is a hot EF task (Lamm, Zelazo, & Lewis, 2006; Mitchell, 2011; Wittmann, Leland, & Paulus, 2007).

Digit Span. Participants completed the Digit Span subtest of the WISC-III (Wechsler, 1991), which includes both Forward and Backward Digit Span. Forward Digit Span required participants to repeat verbatim increasingly longer strings of digits, whereas Backwards Digit Span required participants to repeat the strings of digits in reverse order. Higher scores indicate better performance. This measure of working memory is considered a cool EF task.

The Go–Nogo Task. An adapted version of a previously developed Go–Nogo task was used for this study (Stieben et al., 2007), and yielded measures of both cool and hot inhibitory control. Participants were shown a series of letters and required to press a button as fast as possible whenever a letter appeared on the screen (the go condition) and withhold response whenever a letter was repeated a second time in succession (the nogo condition). In order to provide the same level of challenge for all participants at all ages a dynamic adjustment of the stimulus time was used in the task that adjusted stimulus duration on the basis of performance. Points were added for correct nogo responses and deducted for response errors on both go and nogo trials. Error feedback was provided by a red bar in the middle of the screen following incorrect responses, omitted responses, and late responses.

Children were presented with a practice block followed by three blocks of trials in a fixed order (blocks A, B, and C). In blocks A and C, children gained points quite steadily. These blocks were structurally identical, each consisting of 200 trials, including 66 nogo trials, in pseudorandom sequence. In
block B, children immediately began losing all (or almost all) points due to a change in the point-adjustment algorithm and a reduction in stimulus times. The loss of points was intended to induce negative emotion, such as anxiety and/or frustration. To limit the intensity and duration of children’s distress, block B consisted of only 150 trials, including 40 nogo trials. Although block A provided a relatively cool measure of inhibitory control, Blocks B and C were hot measures of inhibitory control. There is also evidence that, especially in older children, the C block showed increased neural activation, suggesting more mental effort is exerted (Liu, Woltering, & Lewis, 2014). Children were reminded at the beginning of the task, and at the onset of each block, that a high number of points were required to win the ‘big prize’.

Psychophysiological measurements

Electrocardiogram (ECG) data were acquired during the Go-Nogo task at a sampling rate of 1000 Hz using a BIOPAC MP150 system (Biopac Systems Inc., Goleta, CA), with electrodes positioned diagonally across the heart in a standard Lead-II configuration. Data analysis was conducted using ANSLab software (Wilhelm, Grossman, & Roth, 1999). After automatic R-wave detection, each file was visually inspected for artifacts. Small artifacts were corrected when possible by adding or deleting R-wave markers. Files that contained large artifacts or smaller artifacts that could not be corrected were excluded from further analysis. This was the case for two participants. The R-R interval series were then exported and converted into beats-per-minute (bpm). Heart rate reactivity was taken as the difference between the heart rate during the first minute of block B (the emotion induction block) and the last minute (when participants were typically losing all their points and are most upset). Heart rates were not found to differ between groups at the beginning of block B (p > .20).

Emotion scales

At the end of the Go–Nogo task, independent of performance, all children were told that they would receive the big prize and a self-report emotion induction check was administered. Children were asked to rate, on a 10-point Likert scale, the intensity of each of five different emotions (upset, mad, nervous, satisfied, and excited) for each of the blocks. Because they did not differ statistically for the sample as a whole, the three negative (upset, mad, nervous) and positive (satisfied, excited) emotions were summed separately and analyzed as two global measures of experienced negative and positive affect.

For the negative emotions, a repeated-measures ANOVA showed a quadratic main effect of block, F(2,44) = 26.33, p < .001, η² = .54. Planned contrasts revealed that block B was rated as significantly more negative than block A or C (p < .001). No differences were found between block A and C (p = .55). A main effect of block was also found for positive emotion, F(2,18) = 7.56, p = .004, η² = .46, with the planned contrasts showing a significant increase in positive emotions in the last block compared to the rest (all, p’s < .05). Figure 1 shows the results for the positive and negative scales for our sample for each block.

These data show that the emotional context varied across conditions on the Go–Nogo. Block A can be seen as a measure of more cool EF in comparison to blocks B and C, which feature a strong emotion induction. Block B is the negative emotion induction block and, as expected, most negative emotions were reported there. In block C, children reported strong positive emotions. This is likely because the algorithm is normalizing after block B and participants are regaining their points. We point out that questions were asked after the task was completed and participants were told that they had won the big prize which may have biased the answers to be more positive. Block C was seen as the emotion regulation block because participants need to down-regulate their lingering negative emotions while performing the Go-Nogo task. The physiological data collected during the emotional Go–Nogo task confirmed that the emotional blocks (B & C) showed significantly higher levels of arousal compared to the more neutral block A for all participants, F(2,288) = 83.94, p < .001, η² = .37. These data support our classification of block A as a cool measure of inhibitory control and blocks B and C as hot measures of inhibitory control.

Statistical analysis

Outlier analyses were performed on each of the variables used in the present study. Data points were considered outliers and removed from further analysis if they were more than three standard deviations from the mean for each task. The number of outliers was different for each variable but never exceeded more than 5% of the entire sample. ANCOVAs with sex and medication status as covariates were run to determine group differences for the main EF analyses (Typically, first, between DBP vs. Comparison and then between MIXED, EXT and Comparison). If covariates were not significant, they were removed from the analyses. One-tailed p-values were reported for hypothesized effects only (e.g. group differences for the hot EF). Greenhouse-Geisser corrected statistics were reported when assumptions of sphericity were violated for the repeated measure ANCOVAs. Partial eta-squared values (η²) were computed to ascertain effect size. According to Vacha-Haase and Thompson (2004), η² = .01 corresponds to a small effect, η² = .10 corresponds to a medium effect, and η² = .25 represents a large effect.

Results

Clinical questionnaire measures

As expected, the DBP group was rated as having more externalizing problems than the comparison group on the CBCL, t(143) = 17.84, p < .001. The DBP group also showed more internalizing problems than the comparison group, t(143) = 8.68, p < .001, reaching clinical levels of impairment (Mean T-score = 65.3).

The EATQ was consistent with the CBCL as the DBP group had higher scores on Aggression,
Executive function in children with behavior problems  

Effect of subtype was found on Digit Span, \( F(2, 137) = 1.294, p = .28 \). Performance on the Stroop task also did not show any significant difference between the DBP and comparison group as determined by a one-way ANOVA, \( F(1, 135) = 2.33, p = .13 \), and there was no effect for subtype, \( F(2, 134) = 1.61, p = .20 \). Go–Nogo in the cool block, block A, did not show a significant group, \( F(1, 144) = 2.05, p = .15 \), or subtype effect, \( F(2, 143) = 1.10, p = .34 \), for go-accuracy. Nogo accuracy showed a similar null-effect, with no differences between the DBP and comparison group, \( F(1, 145) = 2.89, p = .09 \), as well as the subtypes, \( F(2, 144) = 1.628, p = .20 \). These data suggest that our measures of cool EF do not differ strongly between our DBP and comparison group. Taken together, as predicted, no strong evidence was found that our sample of children with disruptive problem behaviors differed in their cool EF skills from children in the comparison group.

**Hot EF.** As predicted, our measures of affective decision making, the IGT and Delay Discounting tasks, showed group differences. The DBP group performed worse on the IGT, \( F(1, 134) = 3.07, p = .041 \), \( \eta^2 = .02 \) and the Delay Discounting task, \( F(1, 134) = 2.88, p = .048 \), \( \eta^2 = .035 \). The analyses testing for subtype did not yield a significant difference (all \( p's > .11 \)). These results suggest hot EF may be particularly impaired in samples with DBP behavior. Figure 2 shows the bar graphs for the measures of cool and hot EF.

Differences between the DBP and comparison group became visible for the emotional blocks of the Go–Nogo task in contrast to the neutral block A.

---

**Executive function measures**

**Cool EF.** Digit Span showed no group differences as determined by an ANCOVA covarying out effects of medication, \( F(1, 138) = .671, p = .41 \). Similarly, no effect of subtype was found on Digit Span, \( F(2, 137) = 1.294, p = .28 \). Performance on the Stroop task also did not show any significant difference between the DBP and comparison groups as determined by a one-way ANOVA, \( F(1, 135) = 2.33, p = .13 \), and there was no effect for subtype, \( F(2, 134) = 1.61, p = .20 \). Go–Nogo in the cool block, block A, did not show a significant group, \( F(1, 144) = 2.05, p = .15 \), or subtype effect, \( F(2, 143) = 1.10, p = .34 \), for go-accuracy. Nogo accuracy showed a similar null-effect, with no differences between the DBP and comparison group, \( F(1, 145) = 2.89, p = .09 \), as well as the subtypes, \( F(2, 144) = 1.628, p = .20 \). These data suggest that our measures of cool EF do not differ strongly between our DBP and comparison group. Taken together, as predicted, no strong evidence was found that our sample of children with disruptive problem behaviors differed in their cool EF skills from children in the comparison group.

**Hot EF.** As predicted, our measures of affective decision making, the IGT and Delay Discounting tasks, showed group differences. The DBP group performed worse on the IGT, \( F(1, 134) = 3.07, p = .041 \), \( \eta^2 = .02 \) and the Delay Discounting task, \( F(1, 134) = 2.88, p = .048 \), \( \eta^2 = .035 \). The analyses testing for subtype did not yield a significant difference (all \( p's > .11 \)). These results suggest hot EF may be particularly impaired in samples with DBP behavior. Figure 2 shows the bar graphs for the measures of cool and hot EF.

Differences between the DBP and comparison group became visible for the emotional blocks of the Go–Nogo task in contrast to the neutral block A.

---

Table 2: Means and standard deviation for the CBCL and EATQ Questionnaires for three groups

<table>
<thead>
<tr>
<th></th>
<th>Clinical: Externalizers</th>
<th>Clinical: Mixed</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBCL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Externalizing</td>
<td>67.3 (6.5)</td>
<td>73.3 (4.9)</td>
<td>49.8 (8.9)</td>
</tr>
<tr>
<td>Internalizing</td>
<td>53.7 (5.2)</td>
<td>69.2 (5.4)</td>
<td>52.1 (9.6)</td>
</tr>
<tr>
<td>Total Problems</td>
<td>63.4 (6.8)</td>
<td>73.6 (4.2)</td>
<td>50.8 (10.2)</td>
</tr>
<tr>
<td>EATQ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effortful Control</td>
<td>2.69 (0.59)</td>
<td>2.39 (0.42)</td>
<td>3.32 (0.64)</td>
</tr>
<tr>
<td>Aggression</td>
<td>3.41 (0.64)</td>
<td>3.65 (0.60)</td>
<td>2.43 (0.69)</td>
</tr>
<tr>
<td>Depressive Mood</td>
<td>2.66 (0.69)</td>
<td>3.23 (0.67)</td>
<td>2.31 (0.61)</td>
</tr>
<tr>
<td>Fear</td>
<td>2.95 (0.65)</td>
<td>3.11 (0.66)</td>
<td>2.78 (0.64)</td>
</tr>
<tr>
<td>Frustration</td>
<td>3.78 (0.65)</td>
<td>3.98 (0.55)</td>
<td>2.97 (0.65)</td>
</tr>
<tr>
<td>Shyness</td>
<td>2.40 (0.96)</td>
<td>2.54 (0.77)</td>
<td>2.63 (0.86)</td>
</tr>
<tr>
<td>Surgency</td>
<td>3.41 (0.51)</td>
<td>3.36 (0.64)</td>
<td>3.36 (0.60)</td>
</tr>
<tr>
<td>Affiliation</td>
<td>3.65 (0.46)</td>
<td>3.65 (0.61)</td>
<td>3.79 (0.59)</td>
</tr>
</tbody>
</table>

CBCL, Child Behavior Checklist; EATQ, Early Adolescent Temperament Questionnaire.

© 2015 Association for Child and Adolescent Mental Health.
Group differences were found for go-accuracy in the emotion induction block B, $F(1,143) = 6.39, p = .006, \eta^2 = .04$, and in block C, $F(1,143) = 2.76, p = .049, \eta^2 = .02$, with lower accuracies found for the DBP group. When broken down in subtypes, pairwise comparisons for block B showed that these effects were mostly driven by the EXT ($p = .01$, compared to comparison group) rather than the MIXED group (no difference with comparison group, $p = .33$). No effects of subtype were found for block C.

For nogo accuracy, surprisingly, no group differences were found during block B, $F(1,145) = .94, p = .36$, but they were found during block C, $F(1,145) = 9.97, p = .001, \eta^2 = .06$, with the DBP group showing lower accuracy. No effects of subtype were found for block B, $F(2,144) = 1.73, p = .18$, whereas subtype effects for block C showed that the difference with the comparison group was mostly driven by the MIXED group ($p = .008$) versus the EXT group ($p = .14$).

A correlation table (Table S1) showing the relations between different EF tasks can be found as online supplementary information.

**Psychophysiology**

Heart rate reactivity during block B, the emotion induction block, was taken as a physiological index of the degree of emotional reactivity. As expected, heart rate reactivity was larger for the DBP group compared to their nonclinical peers, $F(1,145) = 6.430, p = .01, \eta^2 = .042$. This showed that the DBP group had a stronger physiological reaction to the emotion induction than the comparison group suggesting fewer emotion regulatory capacities. Indeed, performance measures during the emotional block, which could be taken as a behavioral index of self-regulation (e.g. the ability to perform under duress), showed that increased heart rate reactivity was associated with lower performance, $r(83) = -.31, p = .005$. This relation was specific to the emotional block only. To illustrate, Figure 3 shows the relation between heart rate reactivity and nogo accuracy during the emotion induction block for the DBP group.

When investigating subtypes, there was a main effect of group, $F(2,144) = 4.19, p = .017, \eta^2 = .06$. Pairwise comparisons suggested that the difference in heart rate reactivity between the DBP and comparison group was driven by the MIXED ($p = .014$) compared to the EXT subgroups ($p = .51$). These data suggest that children in the MIXED subgroup have a higher physiological reactivity than the EXT.

**Discussion**

The first aim of this study was to investigate whether the EF processes underlying self-regulation were different between children with DBP and their peers. Although performance scores were generally worse for all EF measures in the DBP group, our study found no significant group differences for cool EF as measured by Digit Span, the Stroop Color-Word task, and the initial (relatively cool) block of the Go–Nogo task. However, as hypothesized, differences were found on all measures of hot EF, including the IGT, Delay Discounting, and the emotional block of the Go–Nogo task. These results are consistent with predictions that children with DBP would be especially impaired in EF during emotional states (e.g., Woltering & Lewis, 2009).

Our second aim was to investigate differences between the EXT and MIXED subtypes. A pattern was found where the resulting accuracy differences during the hot emotion induction block of the emotional Go–Nogo task were mostly driven by the EXT as opposed to the MIXED group. It is possible that inhibitory processes are less impaired in MIXED because this subtype may be associated with an overcontrolled style of self-regulation. Studies in anxious populations show that these children perform similarly, and, at times, even better, than their peers on tasks requiring inhibitory control (Günther, Holtkamp, Jolles, Herpertz-Dahlmann, & Konrad, 2004; Hum, Manassis, & Lewis, 2013).

The physiological data suggested that emotional states were not well-regulated in our DBP sample as evidenced by a higher heart rate reactivity during the emotion induction block. This notion was supported by the negative correlation between physiological reactivity and behavioral performance. Consistent with previous research, higher heart rate reactivity was associated stronger with MIXED compared to EXT status (Siess, Blechert, & Schmitz, 2014; Stein & Rapee, 1999).

In sum, our findings suggest that DBP children with poor self-regulation have particular difficulties with hot EF. They show better EF performance in situations that are relatively unemotional, however, as the additional load of emotional regulation is added,

© 2015 Association for Child and Adolescent Mental Health.
these children’s EF skills fail. The results can have practical implications for diagnosing/assessment of EF difficulties. First, cool EF may be less useful in capturing impairments in children with DBP as they often involve simple cognitive responses in controlled and decontextualized test situations. Demands placed on EF abilities in real-life settings are often much more complex as they involve a variety of different emotional situations. Second, these data show the value of manipulating hot and cool EF demands within the same task. In our emotion induction version of Go–Nogo, as a measure of inhibitory control, anxious and frustrating states changed a measure of cool EF into a measure of hot EF. Last, these data highlight the multifaceted nature of disruptive disorders in that degrees of comorbid anxiety in populations with externalizing problem behavior can have negative and/or protective effects on aspects of cognitive and emotional functioning.

A crucial next step in research on EF, as advocated by Zelazo and Bauer (2013), is the development of performance norms across the entire life span. Moreover, the field would benefit from a systematic examination of how cool and hot EF impairment manifests across different psychopathologies or dimensions of psychological constructs (e.g. anxiety). This study focused on levels of anxiety within this population of children with DBP; however, there is also a theoretical incentive to investigate whether those with proactive versus reactive aggression would show a different etiology and underlying mechanism. Sociocultural influences were also not investigated in this study. Aggression, for example, would not necessarily be maladaptive in cases where it is experimental or embedded into a socially (peer-) accepted context such as a gang (e.g. socialized aggression). Another important consideration in this study was that the sample mostly consisted of boys and, due to reasons of statistical power, we could not draw strong conclusions with regards to sex differences. Future studies could investigate sex differences in EF performance under different levels of ‘heat’ and comorbid anxiety. A limitation to this study was that our groups were not matched on demographic variables such as parental education, living situation, and ethnicity, and it is possible that a (constellation of) these factors could have contributed to the group effects. Last, future studies could explore impairments in hot EF in DBP as potential mediators of treatment success. Reliable assessments of treatment efficacy could benefit from the incorporation of measurements that are able to vary levels of emotionality. Adaptive and realistic simulations through, for example, computer games could lead to promising avenues of more ecologically valid assessments or even treatment (Granic, Lobel, & Engels, 2014).

Individual-difference studies are important because they may lead to measurable markers of behavioral problems that could help diagnosis, predict treatment outcome, and potentially lead to insights promoting the formation of novel hypotheses. Though effect sizes are generally weak to moderate, we can conclude from our data that EF measures, particularly hot EF measures, are relevant in explaining individual variability in children’s ability to self-regulate. Moreover, we found evidence that the subtypes of EXT and MIXED showed a distinct profile of behavioral and physiological differences, suggesting that it is important to consider levels of anxiety when examining disruptive behavior disorders.

Supporting Information
Additional Supporting Information may be found in the online version of this article:
Table S1. Correlation table showing relations between different executive function tasks for the entire sample.

Acknowledgements
Funding for this research was provided by the Canadian Institutes of Health Research (CIHR Grant #81170). Acknowledgments go to Connie Lamm, Jim Stieben, Debra Pepler, and Marc Lewis for their work on the Go–Nogo task and setting up the study in its initial phases. The authors have declared that they have no competing or potential conflicts of interest.

Correspondence
Steven Woltering, Department of Educational Psychology, Texas A&M University, 718B Harrington Tower, TAMU, College Station, TX, USA; Email: swolte@tamu.edu

Key points
- Children with disruptive behavior problems (DBP) are believed to have difficulties with self-regulation.
- Hot and cool executive functions, underlying self-regulation, were measured in children with and without DBP.
- Results showed a pattern whereby hot EF tasks were more sensitive in distinguishing DBP from typically developing children than cool EF.
- These findings argue for the use of hot EF in aiding diagnosis, predicting and tracing treatment outcome, as well as further understanding DBP.


Accepted for publication: 16 April 2015